***ФГБОУ ВО «Кубанский государственный аграрный университет имени И.Т. Трубилина»,***

***Российская Федерация***

**66 Интеллектуальные технологии и представление знаний. Лабораторная 6. Разработка методики риелторской оцеки 2020-12-17 310.48**

Заголовок: Разработка методики риэлторской оценки автомобилей в системе АИДОС-Х

Резюме текста:

1. Введение и контекст:

Лекция посвящена лабораторной работе №6 по дисциплине "Интеллектуальные технологии представления знаний" для группы 412. Тема работы: "Разработка методики риэлторской оценки в среде интеллектуальной системы AIDOS и эксплуатация этой методики". Лекцию ведет профессор Луценко Евгений Вениаминович. Обсуждаются этапы развития системы AIDOS (подготовительный, MS-DOS, Windows, Windows 10).

2. Обновление системы AIDOS и инструментария:

Профессор рассказывает о получении новой версии языка программирования (XBS Professional) от фирмы Alaska Software Inc. бесплатно, что позволило преодолеть ограничения старой версии (размер файла, работа с интернетом, совместимость с Windows 10). С 4 ноября по 13 декабря была разработана и скомпилирована новая версия системы AIDOS на новом языке. Старая библиотека XB2NET для работы с интернетом больше не используется, так как функционал встроен в новый язык. Новая версия доступна для скачивания и требует установки в отдельную папку.

3. Методология риэлторской оценки в AIDOS:

Рассматриваются три основных метода риэлторской оценки: по восстановительной стоимости, по потенциальной прибыли и метод аналогии (сравнительных продаж). Лабораторная работа фокусируется на методе аналогии, реализованном в системе AIDOS. Подчеркивается сложность и трудоемкость традиционной разработки таких методик, требующей экспертных знаний и значительных временных и финансовых затрат (оценка в 400 тыс. руб. за 3 месяца). Система AIDOS позволяет разработать методику на уровне лабораторной работы.

4. Подготовка и ввод данных:

Демонстрируется загрузка лабораторной работы №304 в AIDOS. Исходные данные – это Excel-файл с информацией о 401 подержанном автомобиле (марка, модель, кузов, год выпуска, пробег, объем двигателя, КПП, отделка салона, цена). Задаются параметры для обработки данных: классификационная шкала (ценовая категория, 5 градаций), описательные шкалы (характеристики автомобиля, 9 шкал), адаптивные интервалы для числовых шкал. Данные импортируются и кодируются с помощью справочников системы.

5. Создание и верификация модели:

На основе обучающей выборки (закодированных данных) система AIDOS синтезирует статистические и системно-когнитивные модели (процесс занимает около 36 секунд). Подчеркивается важность проверки достоверности модели перед ее использованием. Модель проверяется в режиме 3.4.

6. Применение AIDOS: Решение задач оценки:

Демонстрируются основные задачи, решаемые с помощью построенной модели:

Идентификация (Прямая задача): Определение ценовой категории автомобиля по его характеристикам (режим 4.1.2). Результаты показывают высокую точность распознавания.

Обратная задача (Дедуктивный вывод): Определение характерных признаков для заданной ценовой категории (режим 3). Например, для минимальной ценовой категории характерны малый объем двигателя, МКПП, большой пробег, старый год выпуска; для дорогих – большой объем, АКПП, кожаный салон, новый год выпуска.

Кластерный анализ (Сходство категорий): Анализ сходства и различия ценовых категорий на основе характеристик машин (режим 3.6). Дендрограмма показывает, что очень дешевые и дешевые машины схожи, дорогие и очень дорогие – тоже схожи, а средняя категория ближе к дешевым.

Анализ влияния признаков (Когнитивные функции): Оценка влияния отдельных характеристик (марка, объем двигателя, цвет и т.д.) на принадлежность к ценовой категории (режим 3.7). Визуализация показывает силу и направление влияния каждого признака.

7. Заключение:

Система AIDOS позволяет эффективно разрабатывать и применять методики оценки на основе аналогии, автоматизируя сложные этапы анализа данных и построения моделей. Следующее занятие – зачет.

Детальная расшифровка текста:

1. Введение и контекст

Которая как бы идет 9:40-11:10, но, поскольку там аудиторию вы искали и так далее. С группой 412.

Лабораторная работа номер шесть. Разработка методики риэлторской оценки в среде интеллектуальной системы AIDOS и эксплуатация этой методики.

Дисциплина: Интеллектуальные технологии представления знаний. Занятие ведет профессор Луценко Евгений Вениаминович.

Вот. Так что вот вам основная информация о том, что у нас за занятие.

И сейчас мы сразу же выходим на мой сайт, и я вам кое-что расскажу вкратце. Времени не очень много, поэтому коротко расскажу.

Я здесь немножко изменил вторую страничку сайта. На ней написал об этапах развития.

Так, что-то я не очень понял, откуда у меня здесь столько всяких картинок. А вы видите картинки в чате, нет, ребята? Я же начал новую запись. А там почему-то когнитивные функции сидят. Ну ладно. Вы видите когнитивные функции в чате, нет? Да. Такие картинки разноцветные. Ну, что-то я не понял, как они там оказались, ну ладно. Это я в прошлой с прошлой группой занятие вел.

Значит, смотрим на мой сайт. Ну и картинку сайта увидели, да? Как титульную видеограмму сайта. Ну сейчас пока только ваш экран. Но когда вы включите, да, видно. Нет, ну вот я сейчас послал в чат ее, эту картинку. Вот, она должна в чате появиться.

Я четыре этапа сформулировал развития системы AIDOS. Первый этап – это такой подготовительный. Второй – эра MS-DOSа, IBM-совместимые компьютеры. А первый этап – это другие типы компьютеров. Третий этап – это эра Windows. И четвертый этап – эра Windows десятки.

2. Обновление системы AIDOS и инструментария

И вот я вам могу вам сказать, что недавно, буквально 4 ноября, мне прислала фирма Alaska Software Incorporation, прислала новую версию языка. Причем сам Стефан Пирсих, руководитель этой фирмы, это сделал, письмо написал. Вот. Я просил их об этом. Значит, что у меня нет возможности это приобрести, потому что 2000 долларов, ой, евро, да плюс НДС. Вот его страничка Стефана. Вот. Ну он написал мне, что они посовещались и решили дать мне.

Я поработал с четвертого по, то есть с 4 ноября по 13 декабря, сделал новую версию на новом языке программирования. Ну, насколько он новый? Он это тот же самый язык, только у меня была версия 2009 года, а это версия 2020 года от 29 сентября.

Вот он написал мне: "Хочу сообщить вам, что мы в компании Alaska Software провели внутреннее согласование и решили предоставить вам не для продажи, не для перепродажи, инструментальное средство, версию XBS Professional, XBS++ Professional, и добавили ее в ваш аккаунт на сайте, да, который у меня был, потому что лицензионное у меня программное обеспечение". Это означает, что вы можете использовать эту профессиональную подписку бесплатно в течение 12 месяцев. То есть это бесплатно означает поддержку бесплатную. А так я могу ее использовать сколько угодно до скончания века.

Вот. И вот я где-то несколько раз пытался установить ее на компьютер с семеркой, со старым языком программирования. Потом понял, что лучше взять новый язык и на новом новый компьютер и на нем поставить десятку и на нем все это с нуля ставить. Что я и сделал.

И потом начал изменять те режимы системы AIDOS, которые связаны с работой с интернетом, потому что в старой версии я использовал библиотеку для работы с интернетом XB2NET. А в новой версии без нее. Да, эта библиотека тоже она стоит денег, и обновление стоит денег. Мне это не удалось ее обновить, 400 долларов нету у меня вот для этой цели. Значит, короче говоря, оно и не надо, потому что в новом языке программирования все эти базовые операции с интернетом все есть, функции, поддерживающие базовые операции.

Я переписал вот эти режимы, работающие с интернетом, и 13 декабря откомпилировал полную версию системы, сборку максимальной сборки. То есть до этого у меня некоторые режимы я удалял из нее, потому что компиляция стала невозможной из-за очень большого размера файлов. Писалось, что нет памяти там и такое вот. И, короче говоря, в новой на новом языке я все это добавил, спокойно все эти ограничения сняты сейчас по размеру системы, откомпилировал все.

И старую версию зафиксировал, финальную версию 14-го числа, разместил ее в облаке на память. А все заменил, все инсталляции, все на своем сайте заменил на уже новые, новую систему. И сейчас, если вы хотите новую систему установить, то надо просто все стереть, что было старое там, это скачать и развернуть. Или можно не стирать, а просто развернуть в новой папочке.

Вот. Различие в чем? В том, что разные библиотеки совершенно. И если, то есть путем обновления это сделать невозможно. Нужно именно новую версию загрузить и с ней работать.

Значит, у меня нормально все работает под десяткой, без всяких настроек в операционной системе. И работает в том числе и файлы обновления старта системы, StartAIDOS. Он тоже не работал в десятке, потому что он сразу обращается к интернету. А теперь это все работает, он проверяет обновления, скачивает их. Но пока в интернете, то есть на сайте на моем, расположена версия от 15-го числа. То есть 13-го числа я откомпилировал систему впервые на новом языке. 14-го зафиксировал предыдущую версию финальную из языка на языке 1.9, Аляска 1.9. И сделал версию, еще немножко там доработал то, что было 13-го числа, довел до ума чуть-чуть. Ну я буду и дальше этим заниматься, буду ее развивать и там некоторые недочеты исправлять, связанные с переходом на другой язык. Вот. И разместил 15-го числа уже версию новую для общего доступа. То есть вы можете ее скачивать и пользоваться. И она у вас будет работать без тех проблем, которые были с 1.9. При этом нужно полностью ее устанавливать в своей папочке, не на ту, которая была поверх, а именно с новой папки.

3. Методология риэлторской оценки в AIDOS

Вот. Ну это такая вот информация, она, на мой взгляд, такая положительная эта информация. Теперь, значит, у нас тема сегодня: разработка методики риэлторской оценки в среде интеллектуальной системы AIDOS.

Для того чтобы эту тему раскрыть, я запускаю систему и сообщаю вам, что эта методика риэлторской оценки является лабораторной работой, которую мы будем сейчас просто смотреть, эту работу лабораторную. Сделал такое приветствие с информацией некоторой, с объявлением о том, что это новая версия.

И дальше мы, собственно, что делаем? Устанавливаем эту лабораторную работу и начинаем ее изучать. Это мы делаем в режиме три, в режиме 1.3. Добавить лабораторную работу. Добавляем лабораторную работу третьего типа, работу 304. И мы это и делаем.

И у нее есть Help у этой лабораторной работы. Она также есть в облаке. Help. В качестве хелпа используется, то есть помощи по этой работе, описание работы, используется статья, которую я написал в 2013 году. Некоторые статьи я пишу фундаментального характера, где новые какие-то знания, разработки, концепции, теории, мысли там, вот такое вот. А иногда пишу статьи, которые имеют очень такой конкретный характер и являются просто описаниями решения кое-каких-то задач. Вот это как раз статья такого рода. Она сознательно я ее писал как таким образом, чтобы можно было ее использовать в качестве описания лабораторной работы.

Вот. И устанавливаем эту лабораторную работу.

Появляется экранная форма вот такая, которая, значит, требует пояснения. Значит, здесь мы видим определенные параметры, заданные по умолчанию. Это типы исходных данных: экселевский файл в старом Экселе. Классификационная шкала одна, со второй по вторую колонок. И описательные шкалы с третьей по одиннадцатую. И адаптивные интервалы, то есть от разного размера интервалы с примерно одинаковым числом наблюдений.

Что это за параметры такие странные? Значит, я вам покажу исходные данные. Это экселевский файл, вот, в котором 401 строка с результатами наблюдений. То есть в качестве наблюдения используется, так сказать, описание машин различных марок и моделей. И в качестве, значит, их характеристик рассматриваются кузов, год выпуска, пробег, объем двигателя, тип коробки передач и отделка салона какая. Такие описательные шкалы. А также сама марка и модель машины.

А в качестве классификационной шкалы, того, что мы хотим на основе этих характеристик определять, мы хотим определять ценовую категорию. Ну здесь указана прямо цена на желтом фоне. А, значит, ценовая категория – это числовые диапазоны, которые на основе этой числовой шкалы сделаны. Сейчас мы будем дальше двигаться и увидим эти числовые диапазоны.

4. Подготовка и ввод данных

Вот. А в общем, сейчас мы переходим на этап ввода данных из этого экселевского файла в систему AIDOS. Вот сейчас они вводятся, конвертируются. И мы видим внутренний калькулятор, на котором показано, что у нас есть одна числовая классификационная шкала и три описательных числовых шкалы и шесть текстовых, текстовых описательных шкал.

И мы выбираем здесь в классификационной шкале пять диапазонов числовых, а в описательных – три. И смотрим на то, какие у нас, какое у нас число наблюдений на числовой диапазон получилось. У нас получилось по 80 наблюдений, 81 наблюдение в одном числовом диапазоне.

Возникает вопрос такой: а почему не одинаковые везде число наблюдений? Значит, я могу вам сказать, что не одинаковое число наблюдений по той причине, по какой, по той же самой причине, по которой, когда две ямы копают три землекопа, то никогда не бывает так, чтобы одну яму, чтобы на каждую яму приходилось по полтора землекопа. На одну яму приходится один землекоп, а на другую – два землекопа. Почему? А потому что полтора землекопа, они копать не будут ничего, как вы понимаете.

5. Создание и верификация модели

Поэтому, значит, мы выходим на создание модели. Значит, здесь то же самое. Число наблюдений у нас это целое число. И когда мы делим число наблюдений на число диапазонов числовых интервалов, то у нас получается, вообще говоря, результат, для разных диапазонов может отличаться на единицу.

Это форма о том, что нам делать дальше. Мы то, что дальше, я сейчас уже вам показал одно, один пункт выполнил. А потом сейчас будем другие пункты выполнять. Но сначала я вам скажу, что мы действуем в соответствии со схемой преобразования данных в информацию, а ее в знания, которые приняты в системе AIDOS. Вот этой схемой мы руководствуемся. Сначала мы вводим данные в систему, это формализация предметной области, формируем справочники классификационных и описательных шкал и градаций, обучающую выборку, которая представляет собой закодированные с помощью этих справочников исходные данные. Потом мы создаем модели статистические и системно-когнитивные, определяем наиболее достоверную модель и в ней решаем задачи. Вот этим мы и руководствуемся. И также это описано в хелпе режима 1.3. Вот эта красная линия – порядок выполнения этой работы и других подобных работ.

Значит, теперь мы, когда данные ввели, мы увидели, что у нас в качестве классификационной шкалы пять числовых диапазонов. 1 5, 2 5, 3 5. Значит, 1 5 – это самые дешевые машины, 2 5 – это дешевые машины, 3 5 – это машины средней ценовой категории, 4 5 – это дорогие машины, и 5 5 – это самые дорогие машины. Это у нас классификационные шкалы и градации. Одна шкала – цена, и пять градаций.

Описательные шкалы смотрим. Значит, у нас девять описательных шкал с суммарным числом градаций 245. Ну я вот так скриншотик один покажу одной шкалы. Вот. Названия этих шкал и классификационных, и описательных соответствуют названиям колоночек в файле экселевском исходных данных.

Смотрим на обучающую выборку. Это сами исходные данные, закодированные с помощью этих справочников. Ее можно увидеть в двух видах: в трехоконном интерфейсе, где вверху у нас наименование источника данных, то есть в данном случае это конкретная машина какая-то, хотя здесь название такое несколько обобщенное, то есть можно было бы номер указать машины. Вот. Но в прайс-листе там у них номер не указан, там так было. И в режиме 2.4 мы можем увидеть это же самое, но в форме, похожей больше на экселевский файл исходных данных. Причем похоже, я бы сказал так, не просто похоже, а очень похоже. Сейчас я открою экселевский файл еще раз, и мы их увидим вместе, и вы увидите, что структура абсолютно такая же. Значит, сейчас я ограничу его размера делаю и это делаю ограниченного размера. И мы видим, ребята, смотрите, что структура этих файлов одинаковая. Вот. Можно так вот выровнять даже колоночки, и тогда мы увидим, что она вообще совпадает практически. Так, что-то я здесь не заметил. Так, так, так, так, так, так, так, так, так. Вот. Отделка салона. Цвет. Коробка передач, отделка. Вот сейчас видно, ребят, видно, что хорошо совпадают эти файлы. В этом файле в системе, в экранной форме системы AIDOS, нумерация колонок соответствует реальной нумерации колонок в экселевском файле. И количество тоже их соответствует.

Смотрим дальше. Значит, про сами, про саму задачу давайте я в двух словах вам расскажу. Значит, методика риэлторской оценки – это методика оценки стоимости какого-то объекта продаж. Значит, существует три методики риэлторской оценки основных. Это оценка по восстановительной стоимости, определение восстановительной стоимости, метод аналогии сравнительных продаж и по потенциальной прибыли, которую может дать этот объект продаж.

Восстановительная стоимость. Что это такое? Это сколько стоит создать такой объект. Вот, допустим, стоит дом, и этот дом продается за 3,5 млн. А построить такой дом стоит 2,5 млн. То есть если вы купите строительные материалы, наймете рабочих, проектную организацию со строительными, со строителями реальными, оборудованием необходимым, то они построят такой дом за 2,5 млн. При этом они потратят на это, ну, скажем так, почти год, немножко меньше года. Вот. Вопрос возникает такой: а почему этот дом продается дороже гораздо, чем стоит его построить? Вы обращаетесь к хозяину, они говорят: "Так это ж надо еще построить, а это уже готовый". Это первое. Вот. А время ж деньги. То есть вы будете ждать, когда вам построят, а здесь вы будете пользоваться. Это ж, это же стоит денег? Стоит. Второй момент. Значит, вы деньги заплатили за строительство, а будет ли оно успешно завершено – это еще вопрос. Сейчас же такая обстановка неустойчивая, и курс рубля неустойчивый, и всякие пандемии там и тому подобное. То есть возможно, будет какая-то форс-мажорная ситуация, и не смогут этот дом построить. А здесь уже, пожалуйста, он уже построен точно. То есть когда дом построен, то риски уже исчезают, которые связаны с его постройкой. А когда еще не построен, то существует ряд рисков. Эти риски, значит, тоже стоят денег. То есть когда дом уже стоит реальный, а тот еще, так сказать, неизвестно, будет ли он стоять или не будет через пол, через год, то это тоже стоит денег. Вот эти как раз вот деньги, связанные со временем и связанные с рисками, и есть как раз добавленная стоимость к восстановительной стоимости. Поэтому реальный дом, он может стоить несколько дороже, чем восстановительная стоимость по этим причинам.

Метод потенциальной прибыли, оценки потенциальной прибыли. Если я покупаю фирму, к примеру, готовую, работающую, то для чего я это делаю? Ну, чтобы зарабатывать. Вот я покупаю какое-то малое предприятие, и почему? Потому что это малое предприятие, оно работает успешно и дает ежемесячно хозяину доход 3 млн. Оно небольшое, вот, но вполне нормальный, приемлемый такой доход. Мне говорят: это предприятие стоит 30 млн на рынке. То есть вот можно его купить за 30 млн. Почему? Ну, потому что вот цена будет такой, хозяин говорит этого предприятия, потому что вы примерно за год его окупите, даже быстрее чуть-чуть. Это первое. Вторая причина: стоимость его капитализации примерно тоже соответствует этой величине. То есть если продать здание, оборудование и землю, на которой оно находится, то это примерно вот будет соответствовать этой сумме. Ну то есть вполне обоснованная величина. То есть оценивается, сколько это будет давать вам прибыль, и, значит, это будет не в убыток хозяину с точки зрения его стоимости его капитализации этой фирмы.

И третий вариант оценки, который я считаю наиболее такой обоснованный, потому что он в духе рыночных отношений, именно вот рыночных. Он заключается в том, что метод сравнительных продаж, метод аналогии. В том, что мы должны найти на рынке аналогичные объекты продаж, похожие на тот, который мы хотим продать или купить. Похожие. И вы понимаете прекрасно, что ваш объект продаж, поскольку он похож на другие некоторые, то он может стоить примерно столько же, сколько эти вот другие, похожие на него, аналогичные, в этом же месте и в это же время, что существенный момент, существенное уточнение. В этом же месте, в это же время.

Значит, я могу вам сказать, что существуют методики разработки методик риэлторской оценки. Методики разработки методик. Вот я открываю вам сайт и показываю такую методику. Именно методика разработки методик. И, значит, я могу сразу же сказать, первое впечатление, которое возникает, когда на это смотришь. Вот я вам просто пролистаю вот так. Вот. Ну, все это вызывает уныние, ребята. Вот когда вот так смотришь вот на это все, это вызывает уныние у меня лично. Почему? Да как-то вот неохота вот это все читать, честно сказать, и потом это еще и делать же надо будет. То есть, если так подвести итог, резюме сформулировать просто кратко и четко, то это значит что? Что методика разработки риэлторской методики риэлторской оценки, она довольно-таки наукоемкая, трудоемкая во всех смыслах. То есть это занимает и силы, и время, и, соответственно, имеет стоимость затратам этого этих сил и времени. Причем затраты труда компетентных специалистов, уровня экспертов. И в методиках оценки они должны разбираться на уровне экспертов, и и в самой предметной области. Вот, допустим, мы разрабатываем методику оценки автомобилей. Ну тогда, значит, нам нужны специалисты из фирм Trade-in, которые эту оценку и осуществляют реально. Чтобы они нам все рассказали там, как они все это делают. А мы как экономисты возьмем и разработаем соответствующую методику. Это займет у нас примерно 3 месяца, и мы оцениваем как 400.000 руб. стоимость этой работы. И фирма Акаян, допустим, соглашается, ну они правда своих специалистов там есть, соглашаются на это. Мы заключаем договор и разрабатываем такую методику за это время. А можно взять по-другому поступить. Можно скачать систему AIDOS бесплатно и в ней это разработать эту методику на лабораторной работе, на уровне лабораторной работы.

Я сейчас вам показал, как мы ввели данные. Теперь могу вам сказать что, что сами исходные данные эти взяты с сайта фирмы, которая занимается продажей автомобилей с пробегом. Вот. И я искал сайт, где был бы именно прайс-лист указан. И нашел такой сайт и скачал с него. Адрес этого сайта указан также в этой статье, в списке литературы. Ну вот. Но с тех пор прошло довольно-таки много времени, уже 7 лет прошло. Вот. И многое изменилось. В частности, изменился этот сайт по продаже машин. Теперь у него на этой страничке его этой прайс-листа нет. Уже вообще самого сайта нет. Главную страницу. Вот. Тона Авто Трейд-ин. Вот. Ну, в общем, понятно, что здесь получается. Вот этот сайт, откуда я в свое время скачал прайс этот. Теперь там другая информация.

6. Применение AIDOS: Решение задач оценки

6.1 Идентификация (Прямая задача)

Вот. Значит, теперь мы должны что создать? В соответствии с теми вот методиками, которые я описал, Help и методика обработки данных, информации, знаний, мы должны создать модели и проверить их на достоверность. Это делается в режиме 3.5. Запускаем на исполнение с параметром по умолчанию.

Ребята, извините, это зав кафедрой. Значит, вот так вот оно все время, то начальство звонит непрерывно. Значит, вот у нас модели созданы и проверены на достоверность. Задача решалась уже, скажем так, 36 секунд. Ну та решалась лабораторная элементарные первые, которые мы проходили, на секунды решались, одна, две секунды. Это уже так, видите, почти 40 секунд. Потому что 400 объектов обучающей выборки, уже 400 машин – это уже модель так уже достаточно большой размерности.

6.2 Проверка достоверности модели

Вот. Теперь мы должны посмотреть, насколько эта модель достоверна. Потому что пока мы не узнаем, насколько она достоверна, мы не можем принять решение, можно ли ее использовать для решения задач. Если модель достоверна, тогда можно использовать для решения задач. Если модель недостоверна, тогда ее нельзя использовать для решения задач, это некорректно. И почему? Потому что неправильные будут результаты решения задач. Если вы будете идентифицировать объект, будет неправильная идентификация, будете принять решение, оно будет неправильное решение, то есть не будет достигаться целевое состояние. Будете исследовать моделируемую предметную область путем исследования модели, если она недостоверна, то результаты этого исследования нельзя относить к самому объекту исследования. Понимаете? То есть она не отражает этот объект исследования, поэтому нельзя считать, что если вы модель исследовали, то вы исследовали объект познания, моделирования. Поэтому очень важно всегда проверять достоверность модели. Мы это делаем в режиме 3.4. Это связано вот такой вот ступенечкой с тем, что мы перешли на новый язык, и там есть некоторые различия, и еще я их не отработал.

6.3 Идентификация (продолжение)

Значит, дальше мы решаем сами задачи идентификации. Значит, когда мы проводили синтез и верификацию модели, то они уже были решены. Но сейчас мы выбрали наиболее достоверную модель в качестве текущей в режиме 5-6, и в ней теперь решаем задачу идентификации. Это делается в режиме 4.1.2.

Мы видим сам этот процесс. 400 машин идентифицировано за 3 секунды. Смотрим результаты. В общем, настроение испортилось. Так старался, а тут на тебе. Видим результаты идентификации. Смотрим на нижнее окошко, потому что оно с тем интегральным критерием, который мы по которому модель наиболее достоверна – сумма знаний. Вот этот интегральный критерий – сумма знаний, и результаты идентификации. Смотрим на нижнее окошко и видим, что у нас вот машина, которая первая в списке была исходных данных, в таблице исходных данных, она идентифицирована правильно, отнесена к средней ценовой категории. Есть ложные положительные решения, но они с очень низкими уровнями сходства: 6% и 2%. И вот так, если мы полистаем по разным машинам, то мы увидим, что они очень хорошо все идентифицируются. Ценовая категория очень хорошо определяется этих машин. Вот. И можно вот так вот полистать, посмотреть.

6.4 Обратная задача (Дедуктивный вывод)

Есть обратный разрез, когда мы видим ценовую категорию и все машины, которые к ней относятся. Вот, скажем, машины минимальной стоимости, к ним относятся вот эти конкретные машины. Таким образом, задача идентификации ценовой категории решена по методу аналогии. То есть созданы обобщенные образы ценовых категорий, определены, какие характеристики машин наиболее характерны, не характерны для той или иной ценовой категории. На этой основе определяется ценовая категория, к которой относятся машины.

Теперь рассматриваем решение задачи такое – это обратная задача прогнозирования или обратная задача идентификации, когда мы не по признакам определяем класс, не по характеристикам машины определяем ценовую категорию, а по ценовой категории определяем характеристики машин, которые к ней относятся. Вот, допустим, берем минимальная ценовая категория. Берем модель Inf3, наиболее такая интересная для исследований. И видим, что наиболее характерно для этой ценовой категории, что очень маленький объем двигателя, ткань, отделка салона, коробка передач механика, старая машина. Статья написана в 2013 году, по которой сделана лабораторная работа, и тогда старыми машинами считались машины 96-2007 год. Хэтчбек, похоже, что это Renault, с огромным пробегом. Вот. В этой форме графической та же самая информация, но здесь только наиболее характерные и наиболее нехарактерные характеристики этой машины, такой ценовой категории. Вот в минимальная ценовая категория очень нехарактерна коробка автомат, отделка салона кожа, джип, с огромным объемом двигателя 5,7, новая машина 2013 года на день, на год написания статьи это новые машины были, черного цвета. То есть вы себе сразу представили, что такое дешевая машина, что такое дорогая машина.

Теперь посмотрим на ценовую категорию дорогие машины. Что здесь мы увидим? Мы увидим, что очень большой объем двигателя, джип, кожа салон, автомат, новая, черного цвета. И, скорее всего, это Toyota LC Prado 150 с очень маленьким пробегом. А может быть, и Lexus это, или Volkswagen Touareg, или Toyota LC 200, или Toyota LC Prado 120, а может быть Volkswagen или Infiniti. Это что у нас представляет собой очень дорогая машина. Все это соответствует действительности, все так и есть. Но разные характеристики в разной степени характерны. Наиболее, наиболее характерным таким признаком очень дорогих машин является объем двигателя, то, что кузов джип, и что кожа в салоне, и коробка автомат. Ну а то, что вот он новая, черная и модель – это уже тоже характерно, но в меньшей степени. Так мы видим по каждой ценовой категории. Это мы решили обратную задачу, ребята, идентификации. То есть мы увидели, а что там за классы сформированы у нас? У нас сформировались классы, образы классов ценовых категорий. Вот мы эти ценовые категории увидели сейчас эти классы.

6.5 Кластерный анализ (Сходство категорий)

Теперь мы начинаем решать задачу, это вторая задача в схеме была центральная там она, слева идентификация, потом обратная задача идентификации, принятие решений. И потом сейчас мы задачу решаем исследование моделируемой предметной области. Сюда входят ряд возможностей. Вот, допустим, мы хотим узнать, а что такое Mercedes с точки зрения стоимости? А скорее всего, это будет либо очень дорогая машина, либо средняя ценовой категории. Почему-то даже и очень дорогая, а дорогих нету, то или очень дорогая, или средняя. А Mazda? А Mazda – это либо очень дешевая, либо очень дорогая. А Lexus? Ну Lexus более разумная форма информации. Значит, смотрите, вот Lexus. Значит, либо дорогая машина, либо очень дорогая.

Вот. Смотрим, это как мы можем по какому-то признаку узнать, какая ценовая категория. Вот, допустим, объем двигателя. Возьмем максимальный объем двигателя. О чем это говорит? О том, что это, скорее всего, у нас либо очень дорогая машина, либо с небольшой вероятностью просто дорогая машина. Это очень характерный признак самых дорогих машин.

Вот то, что я сейчас вам рассказываю, ребят, в принципе, эксперты по оценке машин в фирмах, торгующих трейд-ин, например, машинами подержанными с пробегом, они это все знают, конечно. Но вот то, что сейчас мы вам я вам показываю, чем это интересно? Тем, что это мы это видим, во-первых, в количественной форме. То есть здесь у нас не просто качественная оценка: дорого, средне или дешево. А у нас количественно оценивается, насколько каждый из этих признаков влияет на принадлежность машины с таким признаком к этой ценовой категории.

И мы можем увидеть в графической форме эту же информацию в виде традиционном для представления нейрона. То есть центральная часть и рецепторы. Значит, ну, здесь, я надеюсь, вам понравилось, но ничего не понятно. А чтобы было более понятно, нужно уменьшить число рецепторов, потому что здесь модель достаточно большой размерности уже. Ну, значит, мы здесь указываем 32 рецептора и рисуем нейрон с тридцатью двумя рецепторами, которые оказывают наиболее сильное активирующее и тормозящее влияние на этот нейрон, соответствующий этой ценовой категории. Вот, значит, минимальная ценовая категория: очень маленький объем двигателя, ткань, механика, старая машина, хэтчбек, похоже, что это Renault, с очень большим пробегом. Понимаете? Все, ребята, правильно, так и есть.

И мы можем посмотреть это в форме нейронной сети теперь, как это выглядит в этом случае. Ну здесь параметры по умолчанию мы используем, только модель выбираем и рисуем. Значит, здесь мы видим примерно половину модели. Значит, линия связи между значениями показателей, характеристик машины и принадлежностью к ценовой категории, отсортированы все в порядке убывания модуля этой силы связи, то есть без учета знака, силы влияния. И рассортированы слева направо. Сначала идут те, которые наиболее сильное влияние оказывают, потом те, которые послабее, слабее, еще слабее. Что здесь интересного мы отсюда видим? Значит, здесь классы, то есть эти ценовые категории, расположены таким образом, что слева находится та ценовая категория, на которую наиболее сильное влияние, положительное и отрицательное, оказывают характеристики машин. Это что означает, если перевести на русский язык? Что эту категорию легче всего идентифицировать. То есть она очень ярко выражена по тем характеристикам машин, которые относятся к этой ценовой категории, ее легко идентифицировать. Вторая по степени степени детерминированности категория – это категория самые дешевые машины. Тоже она четко выражена по характеристикам. Дешевые машины послабее выражены. Дорогие машины еще слабее, труднее их определить. А самым самым слабо описанным – это средняя ценовая категория. Вот эта, скажем так, все ошибки в основном сводятся к тому, что неправильно определяется средняя ценовая категория, то есть она слабо детерминирована. Нет таких каких-то четких характеристик у этой модели. Но наиболее сильное свойство, наиболее характерное свойство машин средней ценовой категории – это средний объем двигателя. Но и то это слабее намного, чем у других категорий выражено.

Вот. Эта информация ценна была бы для разработчиков методики риэлторской оценки. Теперь смотрим, ребята, насколько сходны эти ценовые категории друг с другом. То есть проводим кластерный конструктивный анализ самих этих категорий, определяем то, насколько они сходны друг с другом по тем характеристикам машин, которые к этим категориям относятся.

Вот мы видим, ребята, что очень похожи друг с другом по характеристикам машины очень дешевые и дешевые. Они прямо очень похожи. А вот дорогие и очень дорогие тоже похожи, но в меньшей степени, чем очень дешевые и дешевые. Причем если очень дешевые и дешевые сходны на 81%, сходство у них, то очень дорогие и дорогие тоже очень сходны, но всего лишь на 70%, не на 81, а на 70. И что интересно у нас относительно средней ценовой категории? Посмотрите, она как бы, скажем так, тяготеет к дешевым машинам и к дорогим машинам. Но она отличается от очень дешевых и от очень дорогих. Она вот именно к тем, которые между ними. То есть между средней категорией и очень дешевыми находятся дешевые машины. Между средней категорией и очень дорогими машинами находятся дорогие машины. Вот она к ним и примыкает. Что совершенно естественно и логично. И убеждает нас в том, что это все вполне разумно то, что я сейчас вам рассказываю. Так, что-то я не тот скриншот поместил. Поместил его два раза этот скриншот. Ну один раз я, наверное, удалю сейчас. Вот.

Это график межкластерных расстояний, ребята. По этой методике расчета и рисования когнитивных дендрограмм, дендрограмм когнитивной кластеризации, написана статья и получено свидетельство Роспатента. Это идет речь об истинной кластеризации, ребята. То есть в отличие от алгоритмов кластеризации, которые в системах статистических используются, где манипуляция с матрицей сходства осуществляется, то когда объединяются классы, то в матрице сходства производятся корректировки. То в системе AIDOS не в матрице сходства производятся корректировки, а в исходных данных. И потом заново формируется модель. То есть там, если что-то объединяется у нас какие-то классы, то они на самом деле объединяются, а не как бы объединяются на основе вот этих вот манипуляций с матрицами сходства. И авторское свидетельство получено на этот алгоритм единой когнитивной кластеризации.

6.6 Анализ влияния признаков (Когнитивные функции)

И теперь, ребята, я покажу вам интересную форму, которая называется когнитивные функции. В этих формах показано, как какой-то, какая-то характеристика машин влияет на их стоимость.

Вот смотрим, как марка влияет на стоимость машины. Вот если марка Renault, то это очень дешевая машина. Мы получаем об этом большой объем информации. Здесь на этой форме изображено количество информации в виде цвета. Без линии соединять бессмысленно. В виде цвета. Значит, красный цвет – это большое положительное количество информации, синий цвет – это большое отрицательное количество информации. То есть красное – это информация о принадлежности машины с таким свойством к этой категории. А синий дает информацию о непринадлежности. Ну и промежуточные цвета спектра, они означают, соответственно, что принадлежность-непринадлежность, но в большей или меньшей степени. Ну, допустим, вот Lexus – это, скорее всего, очень, то есть это машина дорогая. А вот Renault – дешевая. Все-таки я поставлю эти вот линии, потому что они показывают, где максимальное количество информации и минимальное.

Вот смотрим, ребята. Вот у нас машина BMW, наибольшее количество информации о том, что она очень дорогая, исходя из того, какая марка. Infiniti тоже очень дорогая. А вот, допустим, Dodge или там Ford точно не очень дорогая. Либо дешевая, либо средняя, либо очень дешевая, либо дешевая.

И вот мы так можем узнать, как каждый показатель влияет на ценовую категорию. И нарисовать все эти формы. То есть как влияет марка на ценовую категорию, как влияет марка плюс модель. Ну здесь можно сделать, чтобы было видно, но я здесь не делаю это, но можно сделать. Может, сейчас и сделаю. Как влияет год выпуска? То есть машины дешевые – это старые. Как влияет пробег? Чем больше пробег, тем больше похоже, что она будет дешевая. Как влияет объем двигателя? Чем больше объем, тем более дорогая. Все здесь очень хорошо видно. И даже видно, как цвет влияет. Как влияет коробка передач? Отделка салона. Вот, скажем, кожа – это очень признак очень дорогих машин, а ткань – это признак очень дешевых машин.

И давайте я сейчас задам настройку по ширине, задам ну 4К по ширине. Вот. И мы можем посмотреть форму, как влияет марка-модель. Вот сейчас мы уже видим, как она влияет. То есть здесь марка-модель. Марка и градация является модель. Так, это еще не то. То есть марка и градация является марка, да, в шкале марка. А здесь марка-модель, и здесь будет и марка, и модель указана. Вот они. Ну уже здесь видно. Если бы еще не было видно, то тогда можно было бы шрифт уменьшить. Ну и так, в общем, видно. Вот. Смотрим, что у нас там в папочке. В папочке у нас этот файл не сохранен. Сейчас перерисует, и я в чат отправлю. Вот это отправлю. А остальные сейчас перерисую. Но уже не 4К, а разумного, разумной ширины.

7. Заключение

Так, у нас занятие заканчивается через 4 минуты. Так, что-то я как-то много. Ладно. Вот, ребят. Так что всего самого хорошего вам. На следующем занятии зачет будет. До свидания. До свидания. До свидания. До свидания.